Session Goals

1. Define opportunities to use record linkage methods for healthcare research

2. Understand and describe steps to perform data linkage in healthcare

3. Identify barriers and challenges to conducting record linkage

Open the black box of linking electronic health records for research to promote expanded/enhanced use

Conflict of Interest statement

- Neither Dr. Ong nor Dr. Reno have a financial or any other conflict of interest related to this talk.

- We assert that we have no financial interest in any of the social platforms or third party vendors for data mining mentioned in this talk.

Agenda

- Introduction to Record Linkage and CU Record Linkage Planning Workbook

- Workbook Use Cases

- Workbook Feedback

The Promises of Linked Data

- **Between Institutions**: "Fill in the gaps"
  - Many events & outcomes occur outside academic referral centers

- **Within Institutions**: Detect/merge duplicate records
  - 10-15% patients with duplicate records

- **Big Data:** Linking to the larger world
  - Huge array of electronic data

- **Improved data quality**

Definition

Record linkage:

a process to determine if two or more records belong to the same entity
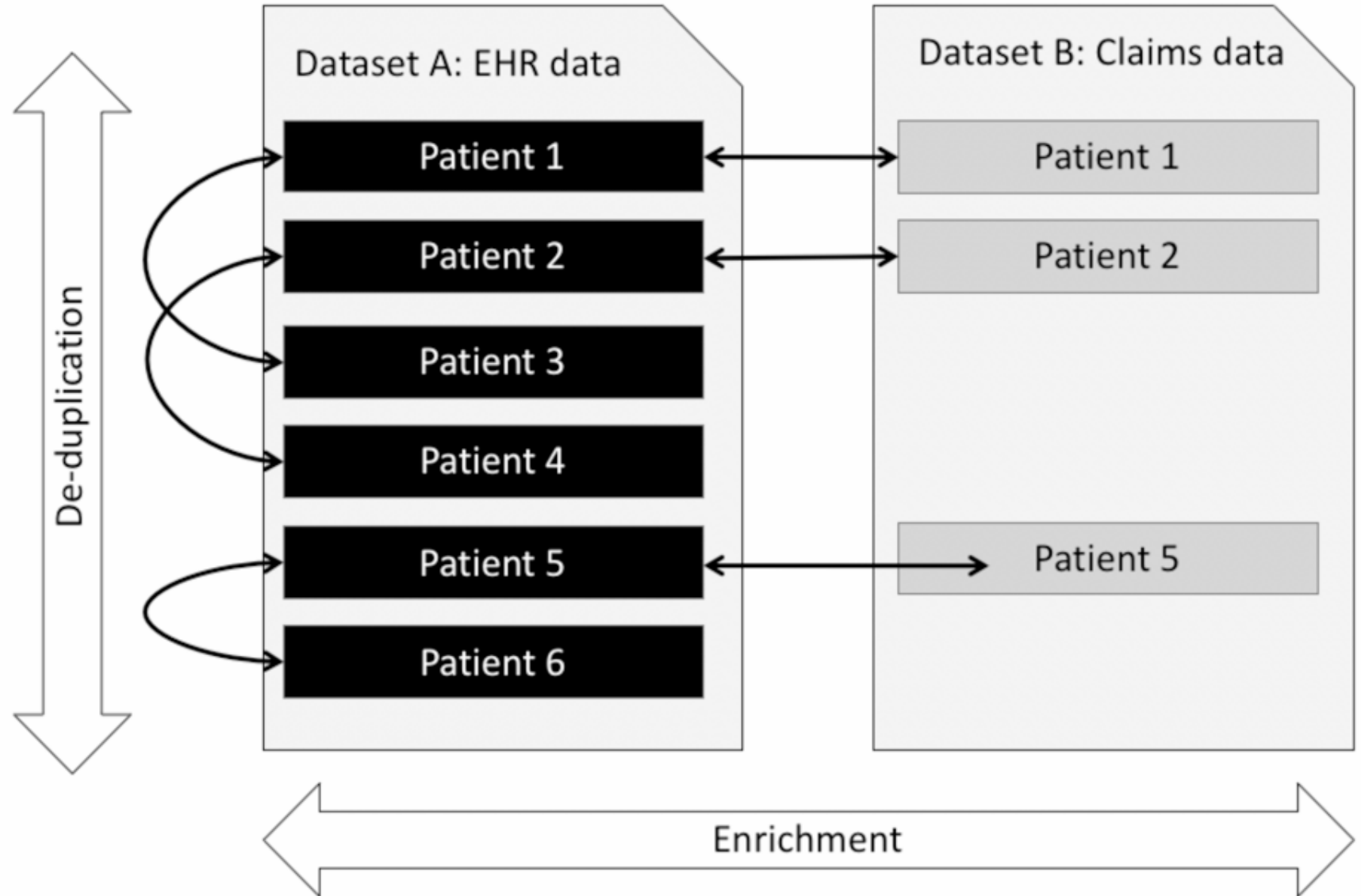
Terminologies:

- Clear Text (CTRL)

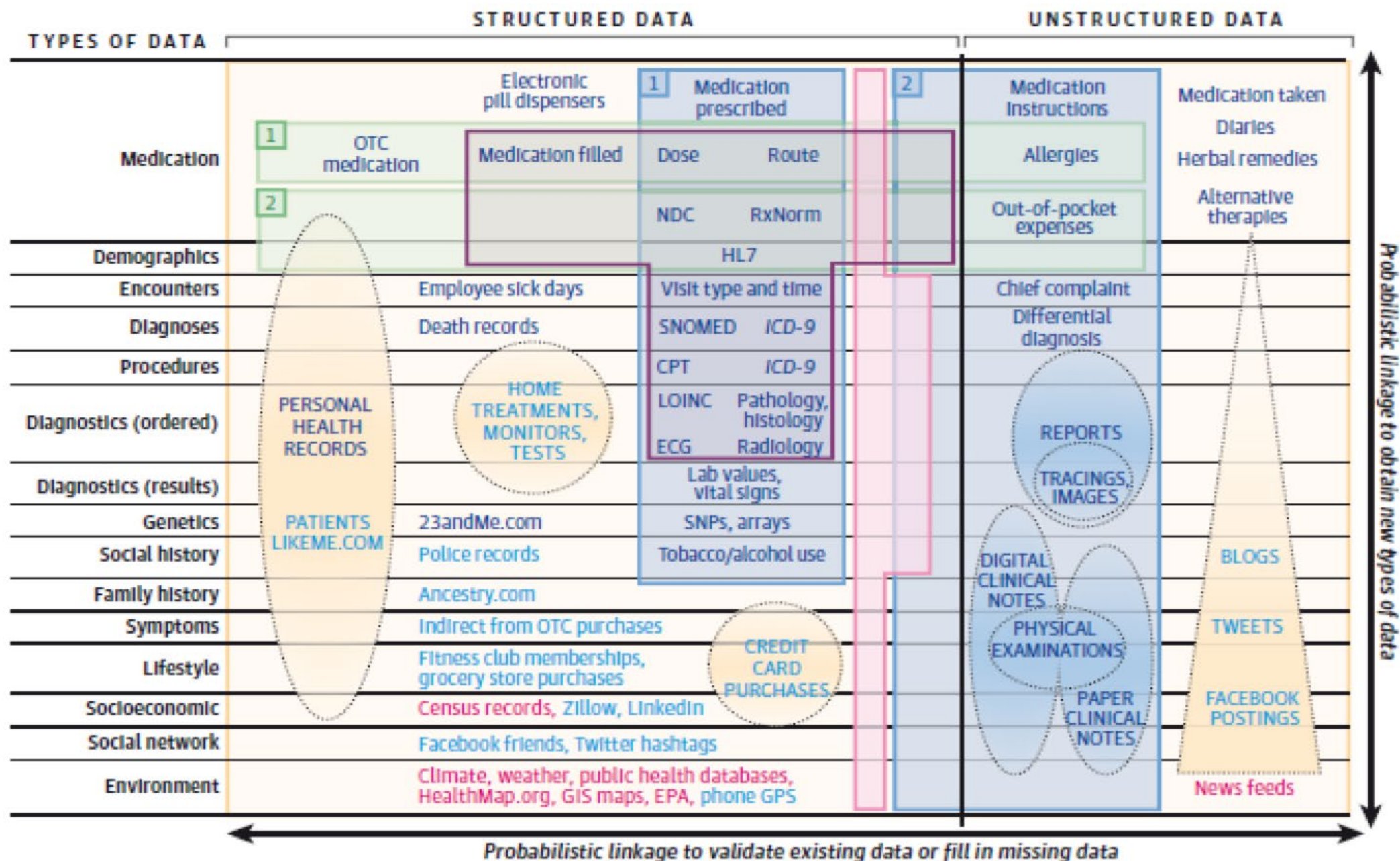- Privacy Preserving (PPRL)

- Incremental Record Linkage

John Smith

ef61a579c907bbed674c0dbcbcf7f7af8f85
1538eef7b8e58c5bee0b8cfdac4a

Objective

Data de-duplication versus data enrichment using record-linkage

STRUCTURED DATA

UNSTRUCTURED DATA

Medication	OTC medication [1] [2]	Electronic pill dispensers / Medication filled	Medication prescribed [1] / Dose / Route / NDC / RxNorm [2]	Medication Instructions / Allergies / Out-of-pocket expenses	Medication taken / Diaries / Herbal remedies / Alternative therapies

HL7

Demographics					
Encounters		Employee sick days	Visit type and time	Chief complaint	
Diagnoses		Death records	SNOMED ICD-9	Differential diagnosis	
Procedures			CPT ICD-9		
Diagnostics (ordered)	PERSONAL HEALTH RECORDS	HOME TREATMENTS, MONITORS, TESTS	LOINC Pathology, histology / ECG Radiology	REPORTS	
Diagnostics (results)			Lab values, vital signs	TRACINGS, IMAGES	
Genetics	PATIENTS LIKEME.COM	23andMe.com	SNPs, arrays		
Social history		Police records	Tobacco/alcohol use	DIGITAL CLINICAL NOTES	BLOGS
Family history		Ancestry.com			
Symptoms		Indirect from OTC purchases		PHYSICAL EXAMINATIONS	TWEETS
Lifestyle		Fitness club memberships, grocery store purchases	CREDIT CARD PURCHASES		
Socioeconomic		Census records, Zillow, LinkedIn		PAPER CLINICAL NOTES	FACEBOOK POSTINGS
Social network		Facebook friends, Twitter hashtags			
Environment		Climate, weather, public health databases, HealthMap.org, GIS maps, EPA, phone GPS			News feeds

Probabilistic linkage to obtain new types of data

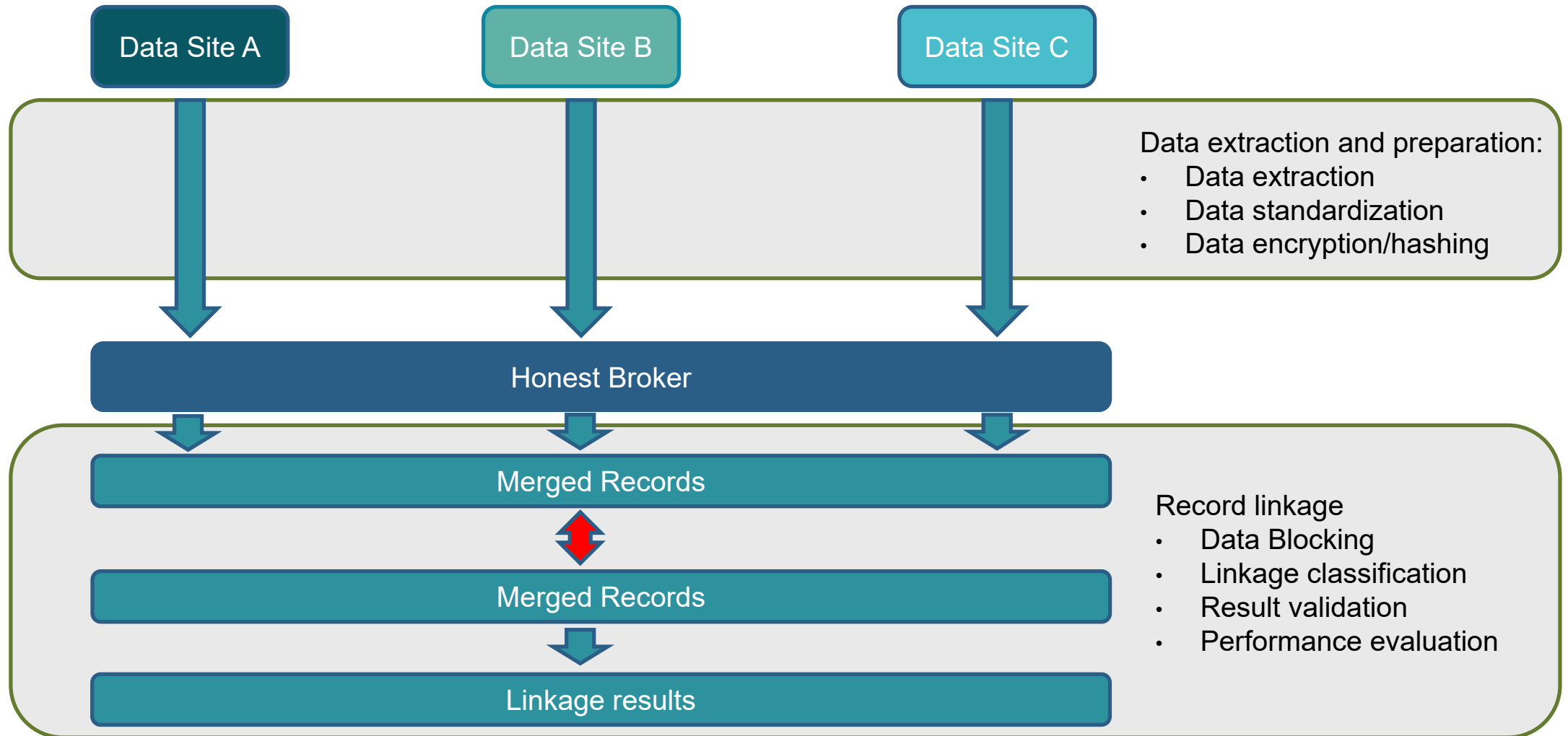*Probabilistic linkage to validate existing data or fill in missing data*

Record linkage scenarios

- A hospital consolidates patient data to ensure that each patient is uniquely identified.

- To understand the relationship between insurance deductibles and readmission rate, a health researcher links hospital data with insurance data.

- A clinical data network estimate number of diabetic patients among its data sites.

Record linkage process

Data Site A   **Data Site B**   **Data Site C**

Data extraction and preparation:
- Data extraction
- Data standardization
- Data encryption/hashing

Honest Broker

Merged Records

Merged Records

Linkage results

Record linkage
- Data Blocking
- Linkage classification
- Result validation
- Performance evaluation

The Challenges of Linked Data

- **No common identifier**
  - Explicitly forbidden in original HIPAA legislation
  - SSN has issues

- **Errors in commonly used variables**
  - Typos (all fields)
  - Marriage/divorce/adoption (Names)
  - Relocation (addresses, phone numbers)
  - Missing values
  - Data formats

Workbook Overview & Use Cases

Funding and Acknowledgements

COPRH Con
Colorado Pragmatic
Research in Health
Conference

ACCORDS

Data Science to Patient Value (D2V)
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

Funding and Acknowledgements

**Data Science to Patient Value
(D2V)**

**Chancellor's Discovery Innovation Fund
(CU Chancellor's Office)**

COPRH Con
Colorado Pragmatic
Research in Health
Conference

ACCORDS

CCTSI